National Centers for Environmental Information
Center for Weather & Climate
Implications for Big Data at NOAA's NCEI

Tim Owen, Chief, CWC Information Services

Transportation Research Board
2016 Summer Workshop
July 27, 2016

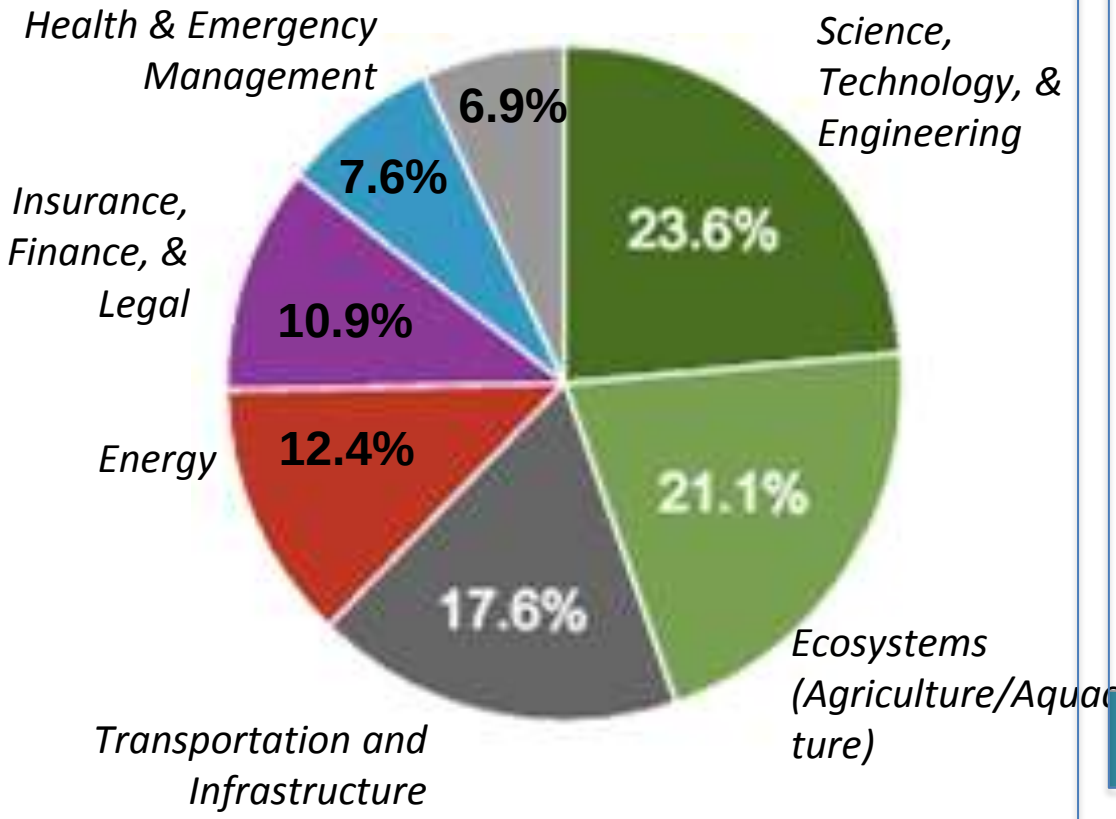
Revealing the Past, Interpreting the Present, and Informing the Future

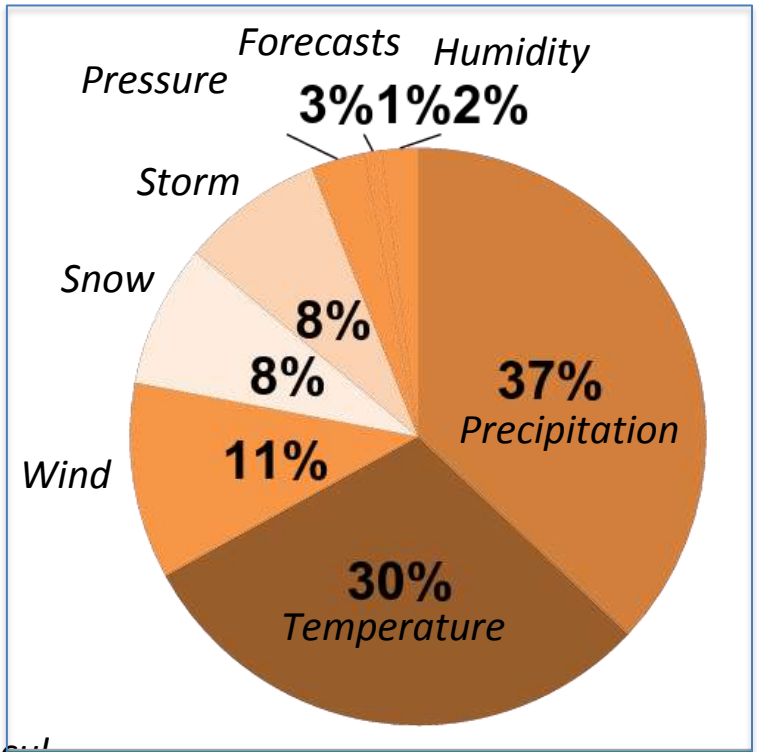
National Centers for Environmental Information

- Responsible for hosting and providing access to one of the most significant archives on Earth, with comprehensive oceanic, atmospheric, and geophysical data
- From the depths of the ocean to the surface of the sun and from million-year-old sediment records to near real-time satellite images
- Nation's leading authority for environmental information

NCEI Users and Requests

By Targeted Sector





By Resolved Theme

Understanding Sectoral Linkages

Linking our knowledge about the changing climate to impacts and decision making

Agriculture

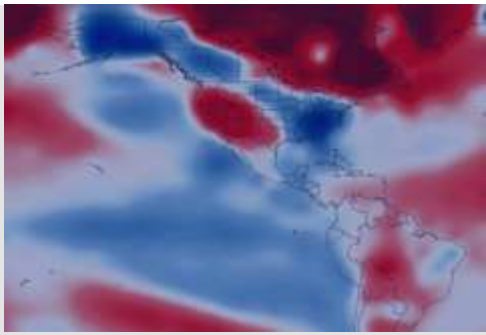
Energy

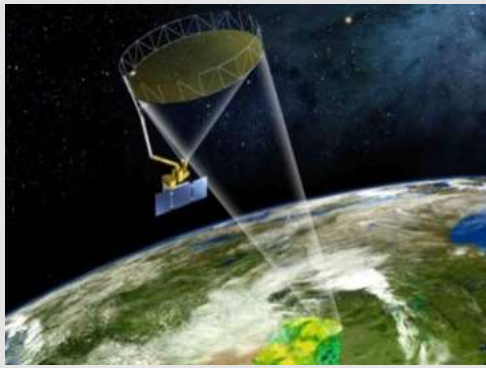
Health

Transportation

Weather and Climate

- World's largest weather and climate data archive with records ranging from paleoclimatology data to centuries-old journals to data less than an hour old
- Data are from land-based weather and climate stations, ships, buoys, weather balloons, radar, satellites, and comprehensive weather and climate models
- A suite of information products is available to describe the national and global climate and to monitor the state of the weather and climate variations, extremes, and trends



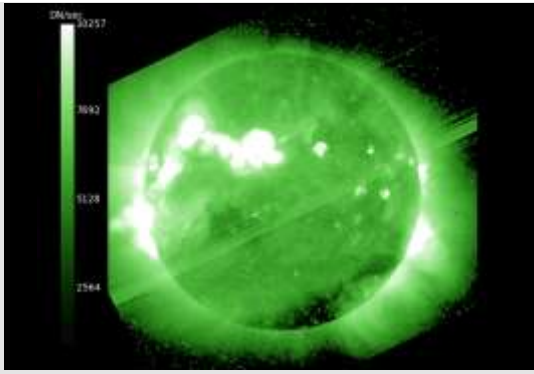


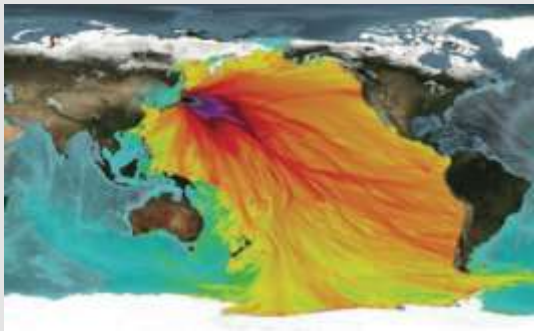
Oceans and Coasts

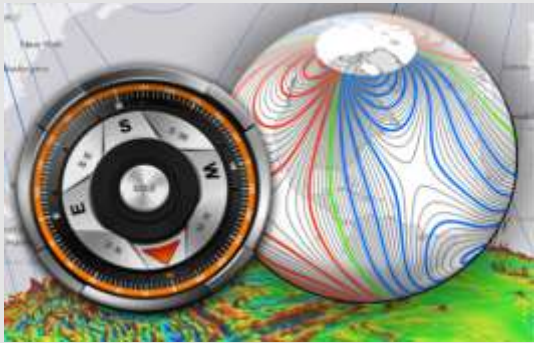
- World's largest archive of oceanographic and coastal data, ranging from water temperatures dating to the late 1700s to present day ocean salinity, nutrients, waves and currents
- Data are from ocean and coast observations, ships, buoys, satellites, remote sensing, and ocean model simulations
- Products and data are used to answer questions about ocean and coast phenomena, management of coastal and marine resources, natural disasters, and marine transportation

Geophysics

- Nation's geophysical data archive ranging from the surface of the sun to the Earth's seafloor and from the solid earth environment to weather in space
- Data are from satellites, space observations, ships, and models provides information on tsunamis, the U.S. Extended Continental Shelf, coastal Digital Elevation Models, geomagnetism, solar, and terrestrial
- Products and data support safe navigation both in space and on our oceans and coasts including the Arctic and the information needed in tsunami forecasting, sea level rise prediction, and storm surge inundation







NCEI products span from local to global, and weekly to decadal scales

	Daily/Weekly	Monthly	Seasonal – Annual	Annual to Decadal
Local	Snowfall Impact Index – FEMA, disaster response	Heating & Cooling Degree Days – Energy Sector	Temperature & Precipitation Outlooks – Agriculture	Coastal Digital Elevation Models (DEM) – Hazard Mitigation
Regional	Hurricane Tracks – Emergency Planners	Solar Activity/Sun Spots – Power Distribution	Billion $ Disasters, Climate Extremes Index – Insurance	Climate Normals – Construction, Infrastructure, Agriculture
National & Global	Tsunami Warning – Emergency Managers	Global and U.S. Climate Summaries – Numerous Sectors	STATE OF THE CLIMATE IN 2013 Annual State of the Climate Reports – Scientists	IPCC & National Climate Assessments – Gov't Policymakers

NCEI and NOAA Have “Big Data”

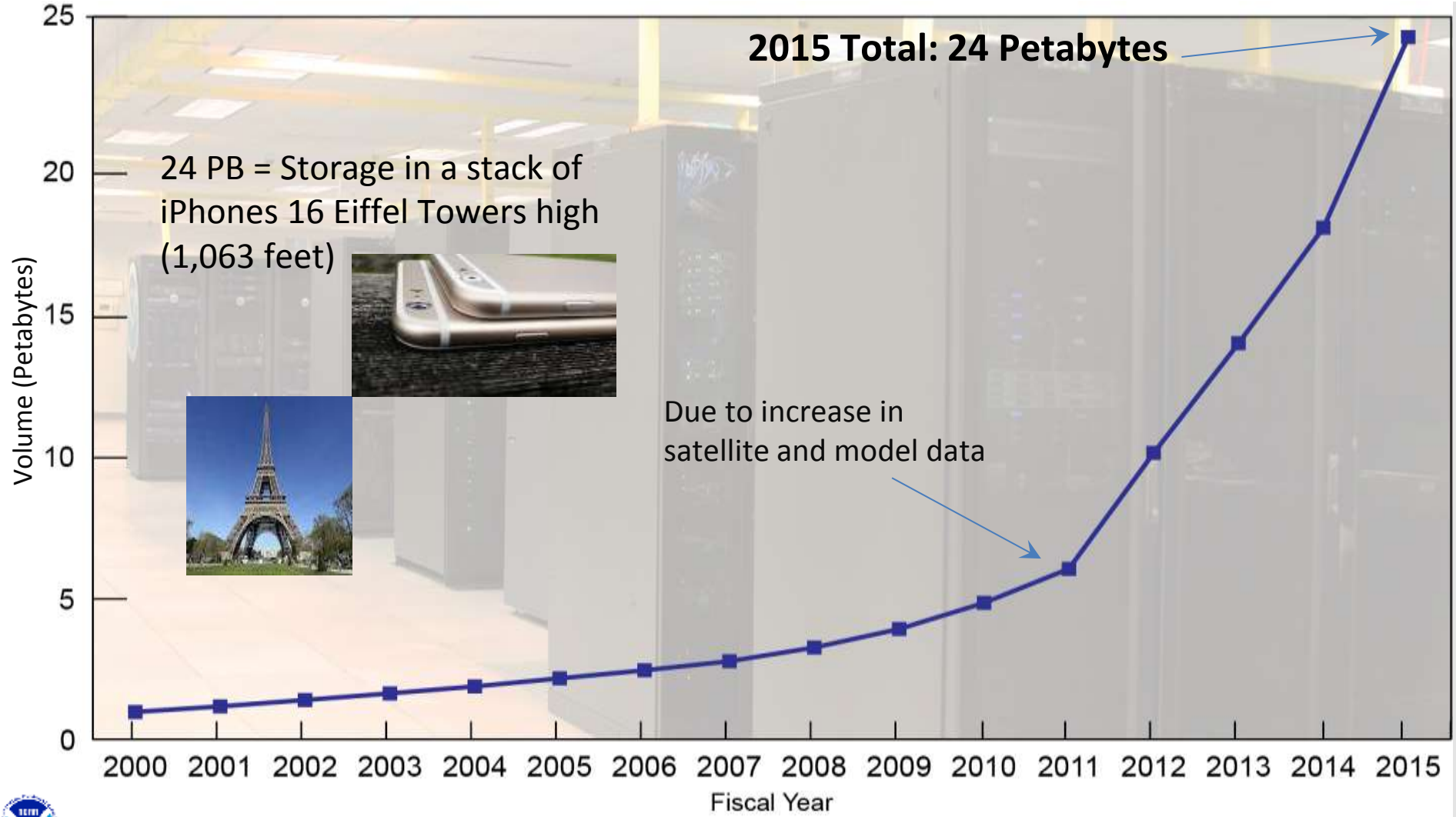
- 10 satellites
- 150+ weather radars
- 3 buoy networks
- 200+ tide gauges
- human observers
- animal telemetry
- 17 ships
- 10 aircraft
- Many numerical models

NOAA data are unique, valuable, irreplaceable, and collected at public expense

(slide adapted from "NOAA 101" briefing)

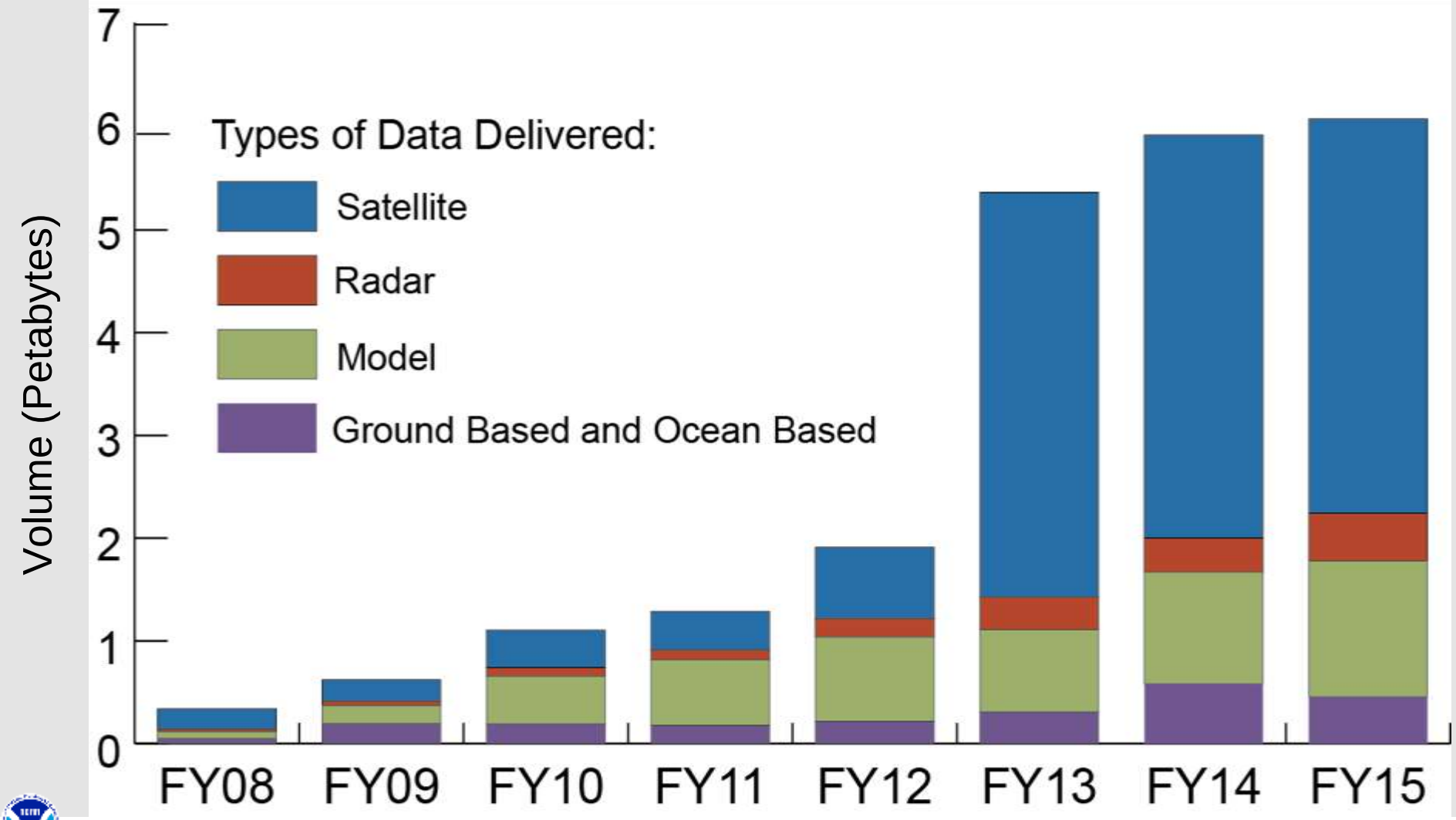
NCEI Environmental Data Archive Volume

Increasing Data Volumes from Station, Model, Radar, and Satellite Sources



NCEI Environmental Data User-Requested Volume

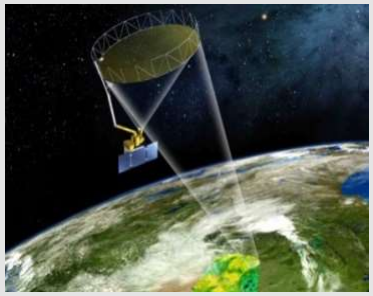
Increasing Data Requests from Station, Model, Radar, and Satellite Sources



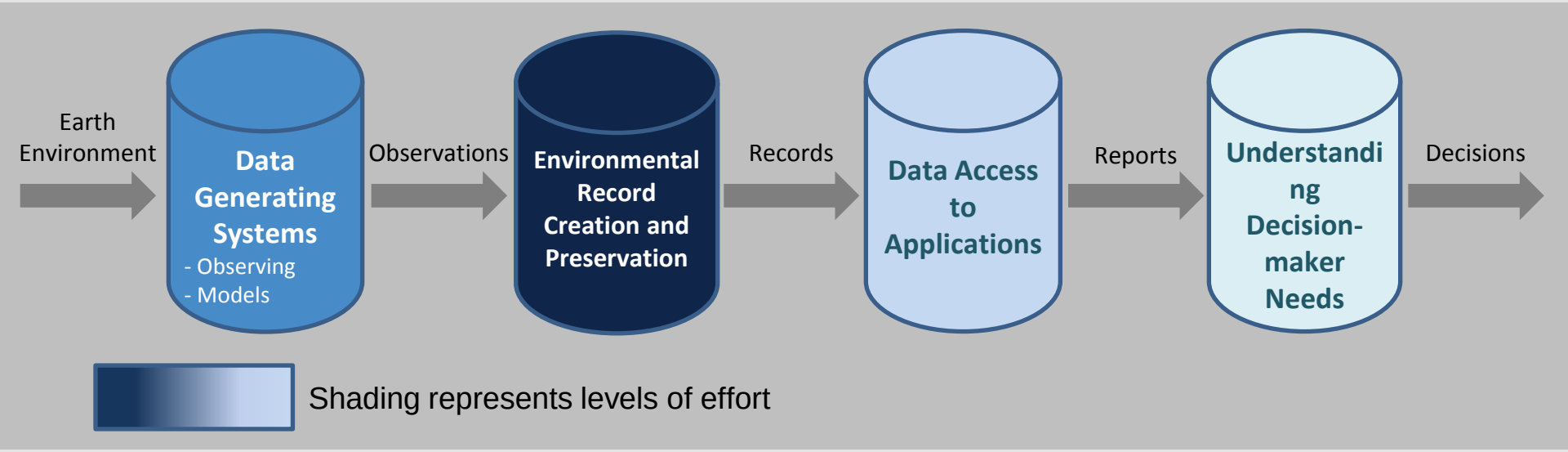
Who Are Our Users? NCEI's General User Profile

Fraction (%)	Typical User	Data or Info Need	Preferred Format	Access Volume	Access Frequency
~70	General business, media, public	Qualitative	Point-and-click, graphics, assessments	Low	High
~15	Researchers, business consultants	Quantitative	Digital downloads	High	Low
~15	Value-added Providers (database scrapers)	Quantitative	Digital downloads machine to machine	Low	High

NCEI Supports the Full Information Lifecycle



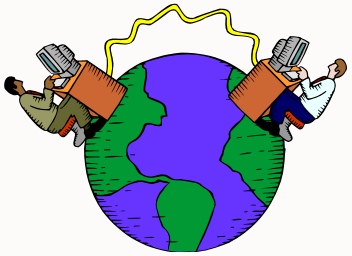
- Makes foundational investments in environmental information production and preservation
- Supports others' application development and policy/decision-making



Vision for NOAA Data Management

All NOAA environmental data are to be:

Discoverable



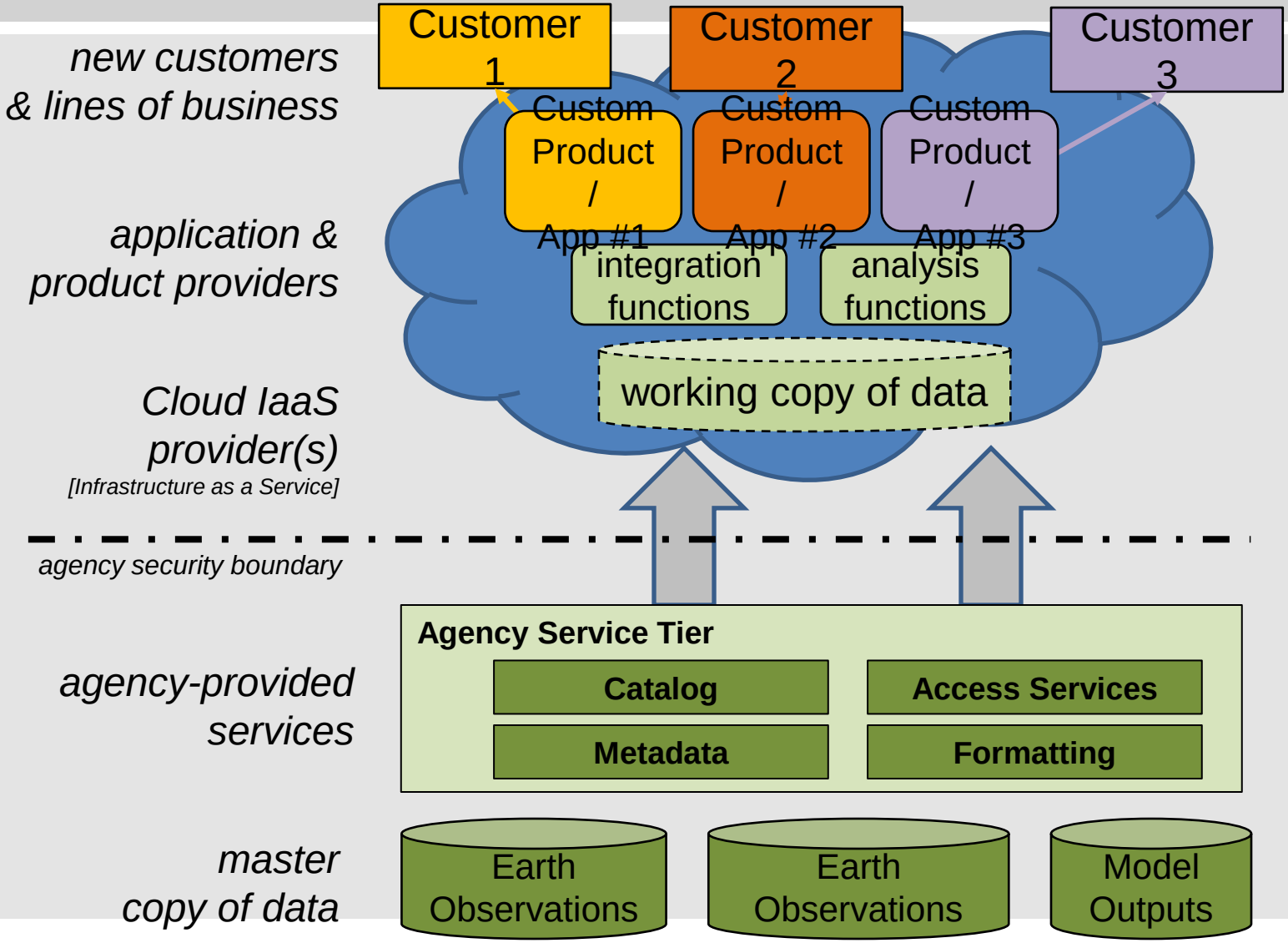
Accessible

Usable

Preserved

for all types of users and applications.

Overview of Big Data at NOAA



Premise of NOAA Big Data Project

- There is additional value in NOAA data that has not yet been realized because of access & infrastructure difficulties.
- If NOAA data were accessible in the Cloud, alongside computing capability, private enterprise might generate new value-added products, services, and lines of business.
- Private enterprise might be willing to support the cost of transferring and storing large datasets because of these new lines of business.
- Self-sustaining partnerships are possible

5 BDP CRADAs announced April 2015

CRADA = Cooperative Research and Development Agreement

Big Data Project - Home |

https://data-alliance.noaa.gov

NOAA NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
UNITED STATES DEPARTMENT OF COMMERCE

» Home » FAQ

NOAA Big Data Project

The Big Data Project is an innovative approach to publishing NOAA's vast data resources and positioning them near cost-efficient high performance computing, analytic, and storage services provided by the private sector. This collaboration combines three powerful resources - NOAA's tremendous volume of high quality environmental data and advanced data products, private industry's vast infrastructure and technical capacity, and the American economy's innovation and energy - to create a sustainable, market-driven ecosystem that lowers the cost barrier to data publication. This project will create a new economic space for growth and job creation while providing the public far greater access to the data created with its tax dollars.

How To Participate

For companies, organizations, and individuals interested in joining with NOAA's Big Data Project, a set of Data Alliances are being formed. Each Data Alliance is anchored by a participating Infrastructure as a Service (IaaS) institution, and represents a market ecosystem consisting of larger companies that represent various economic sectors, such as the weather or insurance industries, specialized small business, value-added resellers, entrepreneurs, researchers and non-profits, etc. The Data Alliance structure allows market forces to act on the identification, extraction, and development of NOAA public data resources, and provides a mechanism for interested parties to work together to develop new business and research opportunities. The organizations comprising the ecosystem built around a particular anchor IaaS provider are free to participate in multiple Data Alliances.

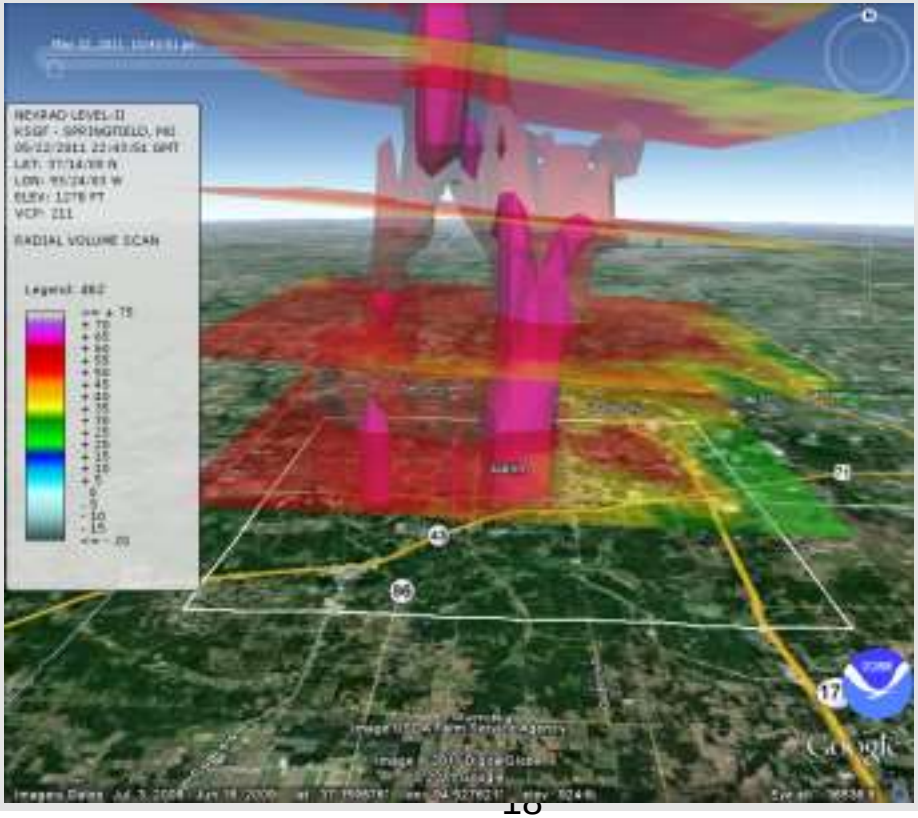
For more information, visit one of the NOAA Big Data Collaborators:

amazon webservices | Google Cloud Platform | IBM | Microsoft | OCC OPEN CLOUD CONSORTIUM

The Big Data Project's press release, issued by the Department of Commerce, can be found at http://www.commerce.gov/news/press-releases/2015/04/us-secretary-commerc...

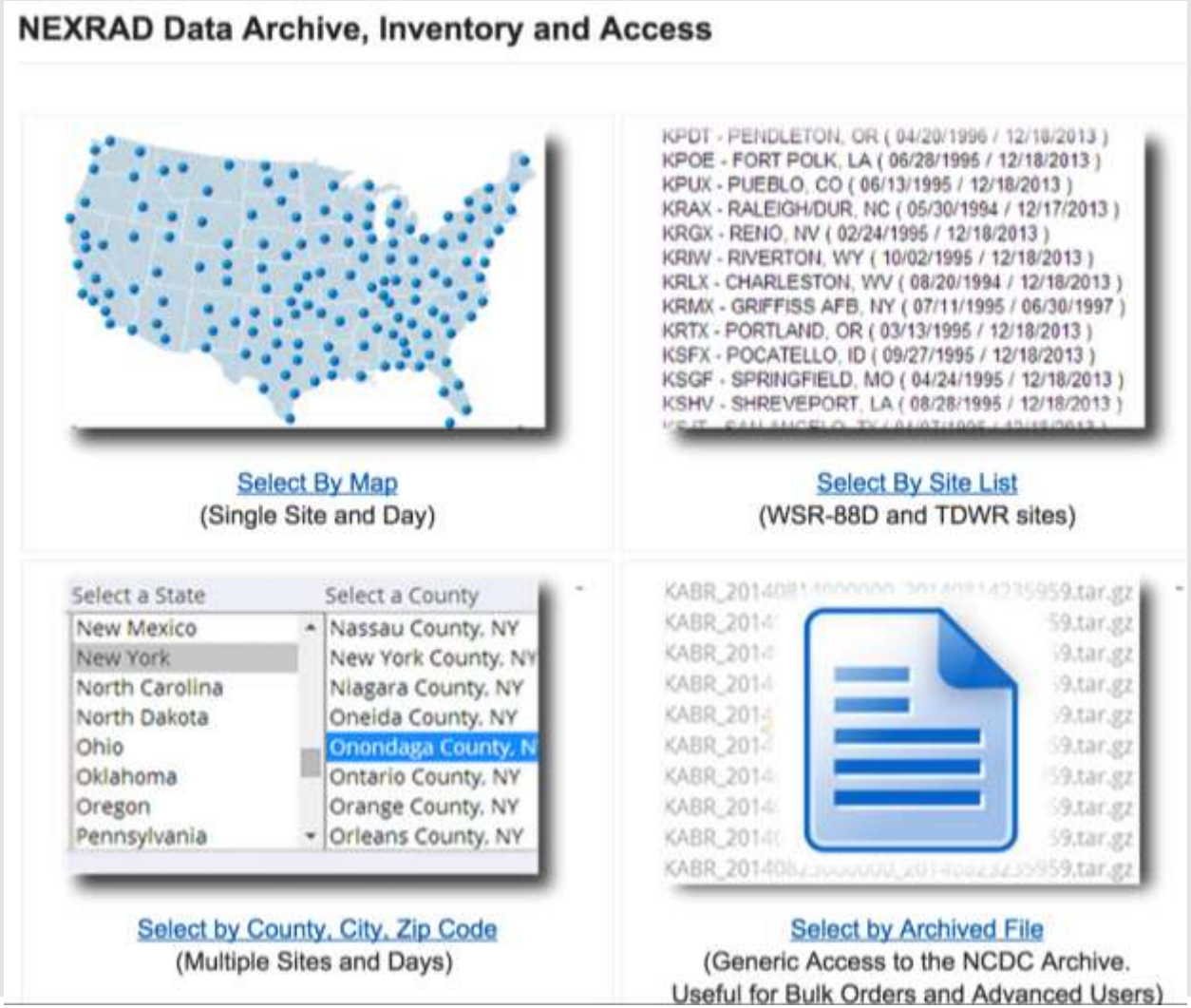
1st BDP Dataset: NEXRAD L2

- NEXRAD = Next-generation Radar
  - Level 2 = reflectivity data from 150+ stations

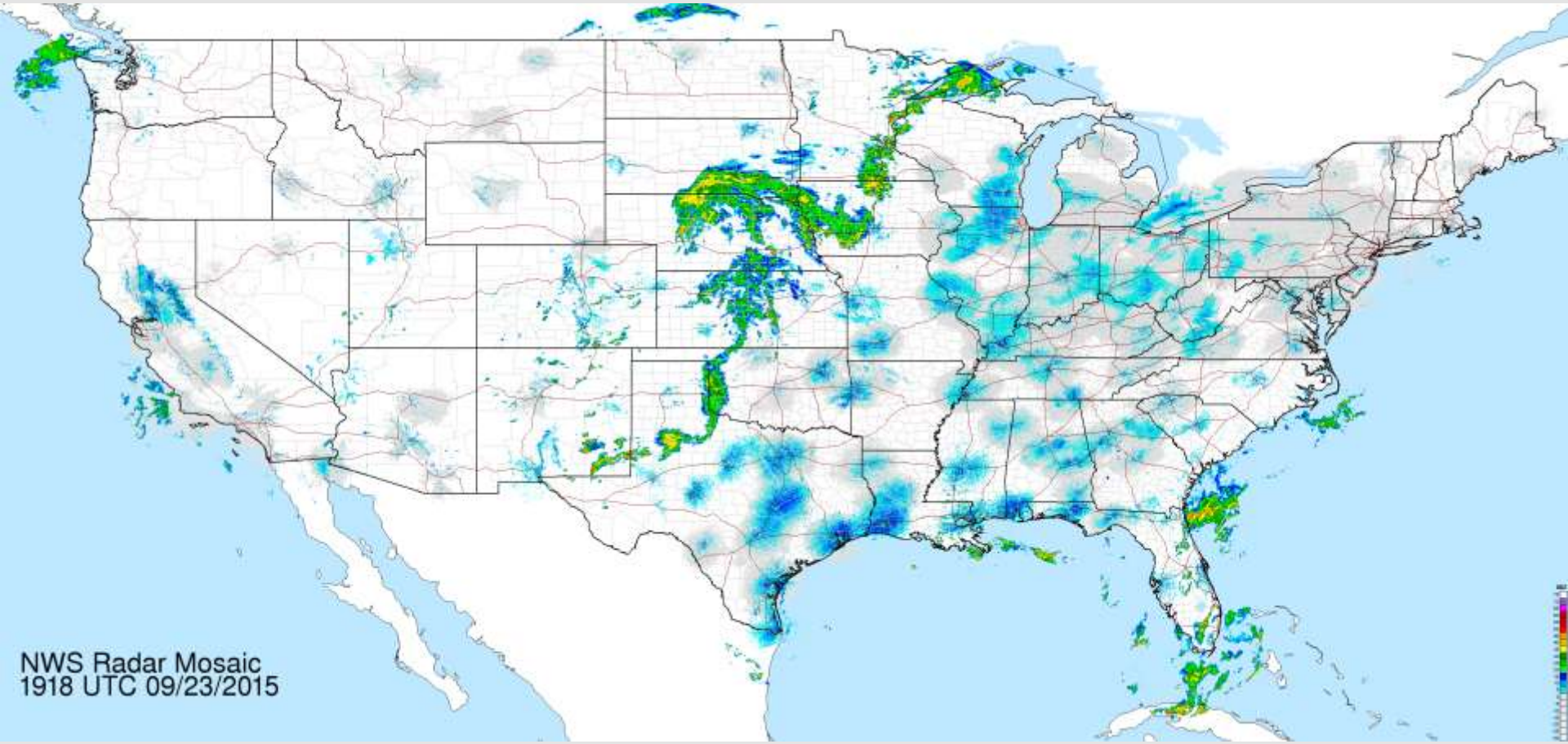


18

1st BDP Dataset: NEXRAD L2



1st BDP Dataset: NEXRAD L2



1st BDP Dataset: NEXRAD L2

- Copying entire archive of Level 2 processed data to Cloud
  - Data from 1991-2015 archived at NCEI.
  - Not “dark” data, but unwieldy in current state
    - Data on near-line tape archive => slow access
  - 270+ TB compressed (over 1PB uncompressed)
  - Transfer to Amazon to be complete in October 2015
    - Other NOAA BDP Cloud partners also interested

- Data can be utilized directly in the Cloud platform
  - Access charges (cost recovery only) may apply
  - CPU charges will apply

NCEI's Information Services

Objective: *Deliver Use-Inspired Environmental Information*

that Supports the Nation's Prosperity and Resilience

Provide Information

Expand and Enrich Use of NCEI's Environmental Information

Understand Users

Understand User Needs and Translate Them Into Requirements

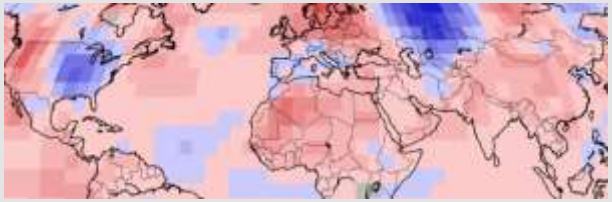
Strengthen Networks

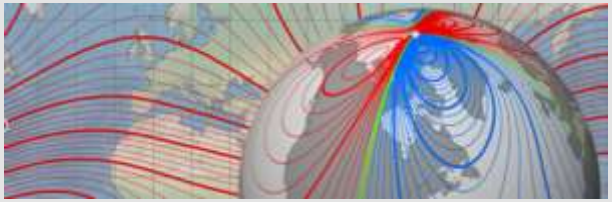
Strengthen Networks for Developing and Delivering NCEI's Products and Services

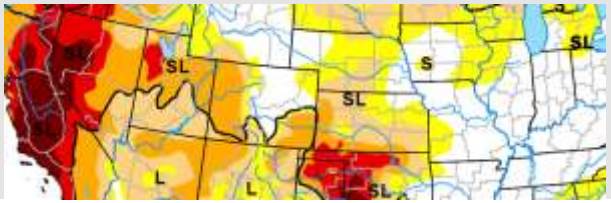
Value of NCEI

Science, Service, Stewardship

- Comprehensive environmental information across economic/environmental sectors and science disciplines for decision making, planning, and operations
- Timely and authoritative data and information with well described origin and certainty
- Service to a wide-range of customers across government, academia and industry
- Preservation of data and information for future generations

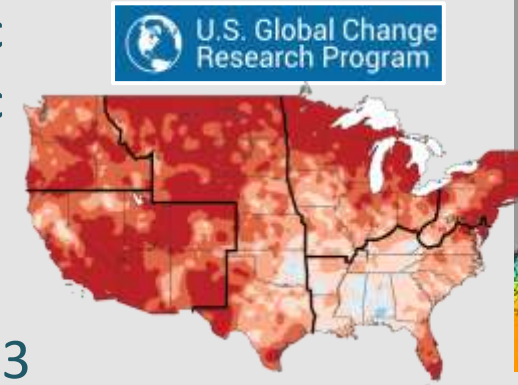


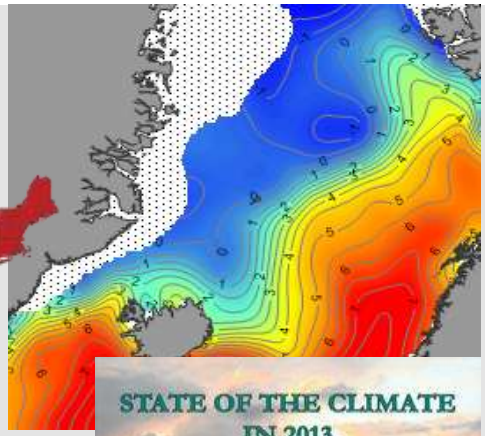


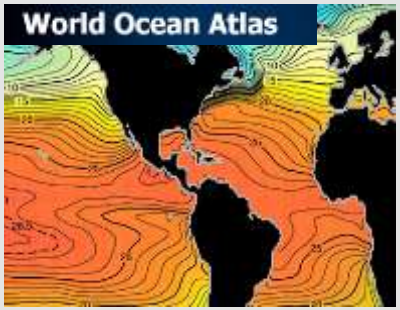


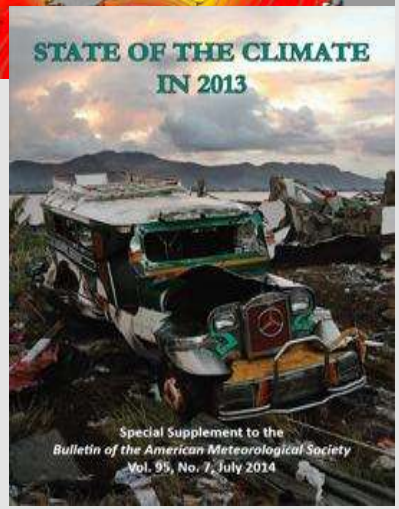
Examples of “Gold” standard NCEI Products

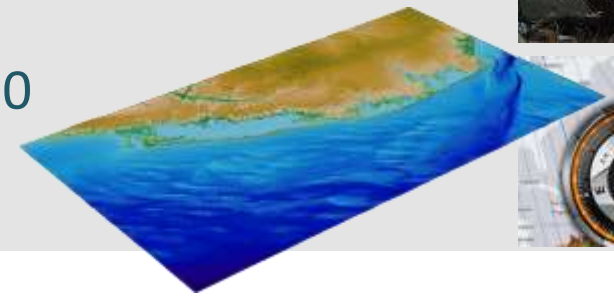
- Climatological Atlas of the Nordic Seas and Northern North Atlantic
- World Ocean Atlas 2013
- National Climate Assessment
- BAMS State of the Climate in 2013
- Explaining Extreme Events of 2013 from a Climate Perspective
- Extended Continental Shelf (ECS) Project
- Post-Sandy Digital Elevation Model
- World Magnetic Model for 2015-2020

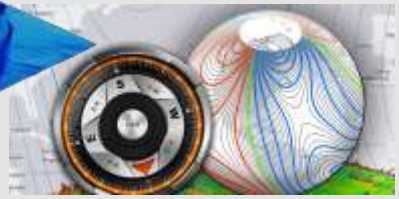








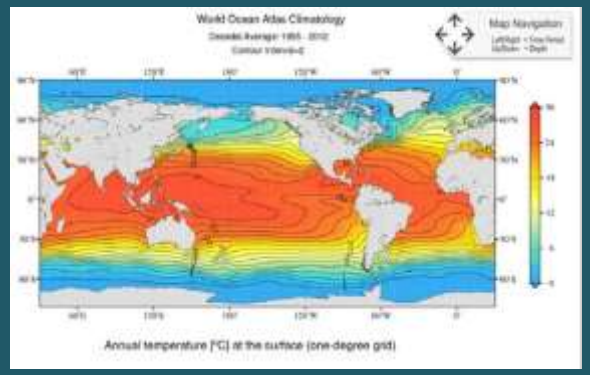




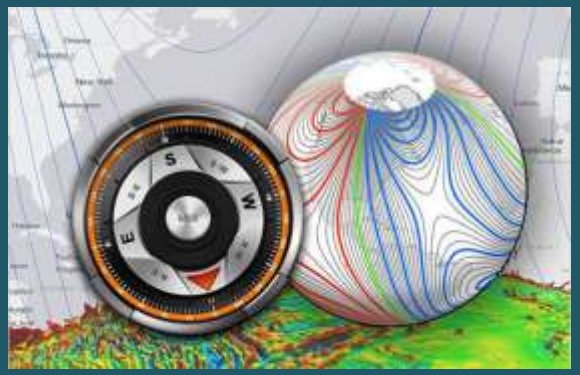
Examples of Reference Products

Climate Normals

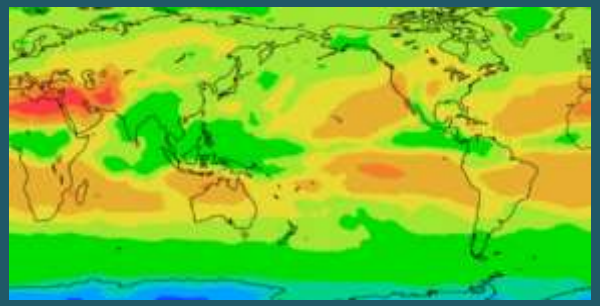
World Ocean Atlas



Earth's Magnetic Field

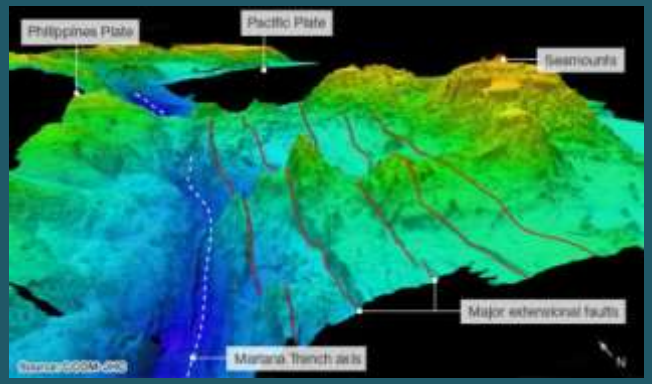


Reference Environmental Data Records



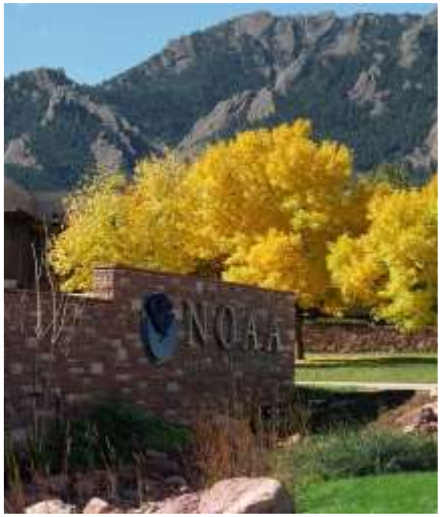
Gulf of Mexico Data

Coastal/Ocean Depths



National Climate Assessment

http://www.globalchange.gov/


NOAA/NESDIS
National Centers for Environmental Information

www.ncei.noaa.gov

NCEI Climate Facebook: http://www.facebook.com/NOAANCEIclimate
NCEI Ocean & Geophysics Facebook: http://www.facebook.com/NOAANCEIoceangeo
NCEI Climate Twitter (@NOAANCEIclimate): http://www.twitter.com/NOAANCEIclimate
NCEI Ocean & Geophysics Twitter (@NOAANCEIocngeo): http://www.twitter.com/NOAANCEIocngeo

BACKUP SLIDES

Tiers of Stewardship

Levels of Data Stewardship

6: National Services and International Leadership
- Lead, coordinate, or implement scientific stewardship activities for a community or across disciplines
- Establish highly specialized levels of data services and product assessments

5: Authoritative Records
- Combine multiple time series into a single, inter-calibrated product
- Establish authoritative quality, uncertainties, and provenance
- Ensure products are fully documented and reproducible

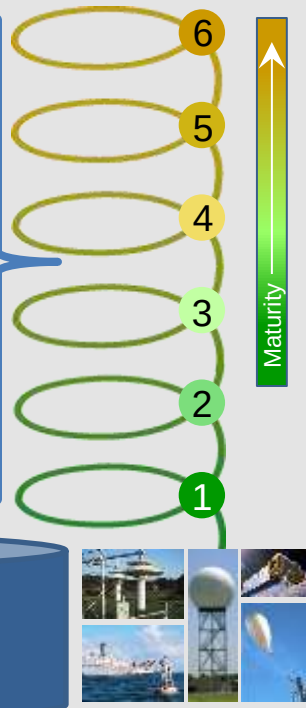
4: Derived Products
- Build upon archived data to create new products that are more broadly useful
- Distill, combine, or analyze products and data to create new or blended scientific data products

3: Scientific Improvements
- Improve data quality or accuracy with scientific quality assessments, controls, warning flags, and corrections
- Reprocess data sets to new, improved versions and distribute to users

2: Enhanced Access and Basic Quality Assurance
- Create complete metadata to enable automated quality assurance and statistic collection
- Provide enhanced data access through specialized software services for users and applications

1: Long Term preservation and Basic Access
- Preserve original data with metadata for discovery and access
- Serve as expert advisors on standards for data providers
- Archive only necessary data using appropriate retention schedules
- Safeguard data over its entire life-cycle
- Coordinate support agreements for sustainable data archiving
- Provide data citation services by mining DOIs

Maturity Matrix Model



How do we Develop & Maintain "Gold Standard" Reference Datasets?

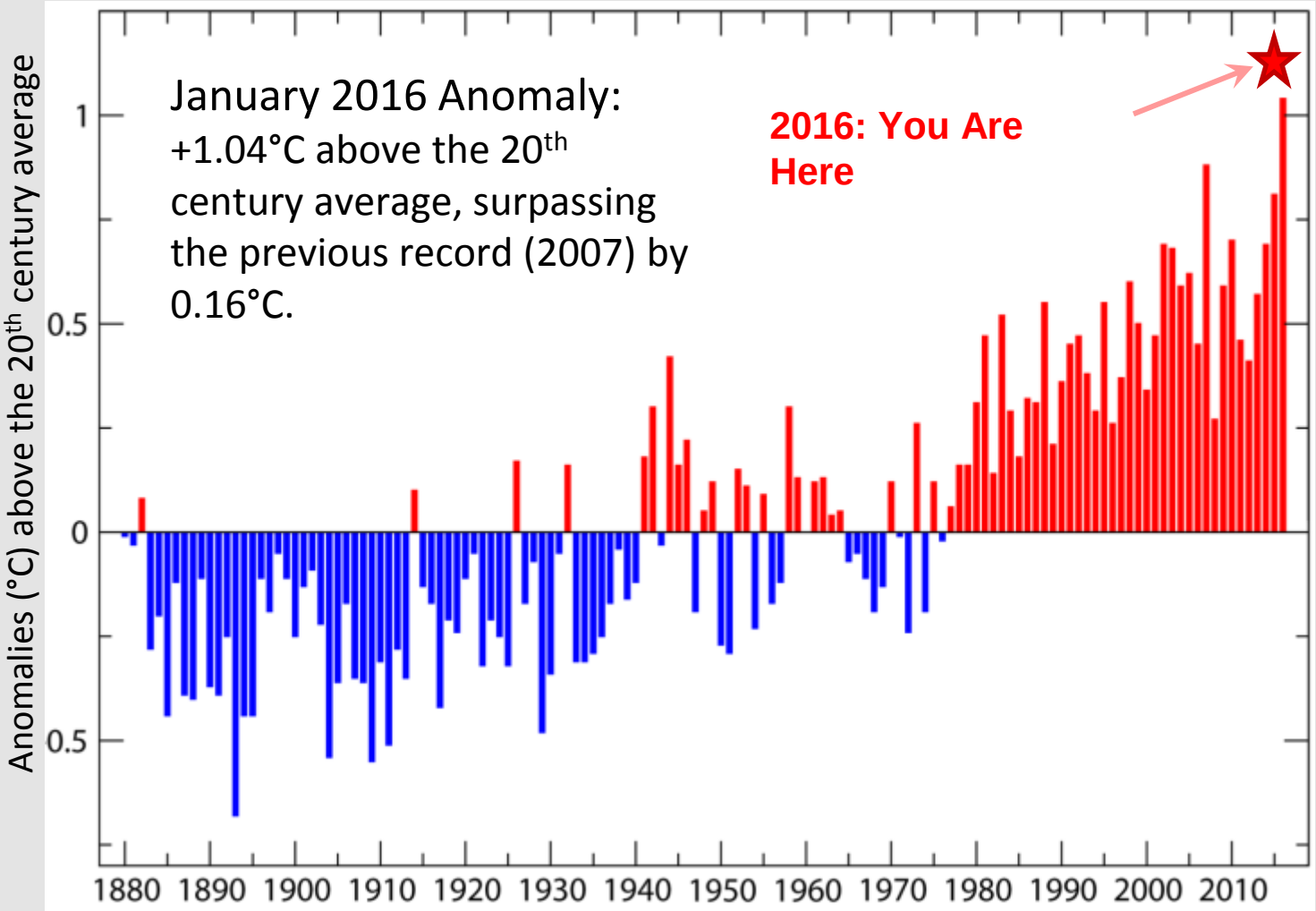
- Apply a maturity model similar to engineering readiness levels - a practical approach
- Goes beyond measurements of quality to *usability, transparency*, and *understanding*
- Assess products using a "Maturity Matrix" covering seven aspects of data maturity

	Level	Data Use	Code Stability	Metadata & QA	Documentation	Validation	Public Release	Science & Applications
Primarily Operations	6	Unified and coherent record; considered scientifically irrefutable following extensive scrutiny	Homogeneous and published error budget	Provenance tracking and reproducibility; meets international standards	Peer-reviewed product algorithm, validation, processing and metadata	Validated independent cross-checks, open inspection, and continuous interrogation	Publicly available from long-term archives	Used in various published applications and assessments
	5	Unified and coherent record	Stable and reproducible	See below	See below	See below	Source code portable and released; uncertainty estimate	See below
	4	Research and Operations	See above	Provenance tracking and reproducibility; meets international standards	Draft Operational Algorithm Description	See below	Source code released; Data available but of unknown accuracy	See below
Primarily Development	3	Research	Minimal changes expected	See below	Peer-reviewed algorithm and product descriptions	Uncertainty estimated over widely distribute times/location; differences understood	See below	Provisionally used in applications and assessments demonstrating positive value
	2	Research	Some changes expected	Research grade (extensive)	CTD	Uncertainty estimated for select locations/time	Data available but of unknown accuracy	Limited or ongoing
	1	Research	Significant changes likely	Incomplete	Draft concept of Theory Document (CTD)	Minimal	Limited data availability	Little or none

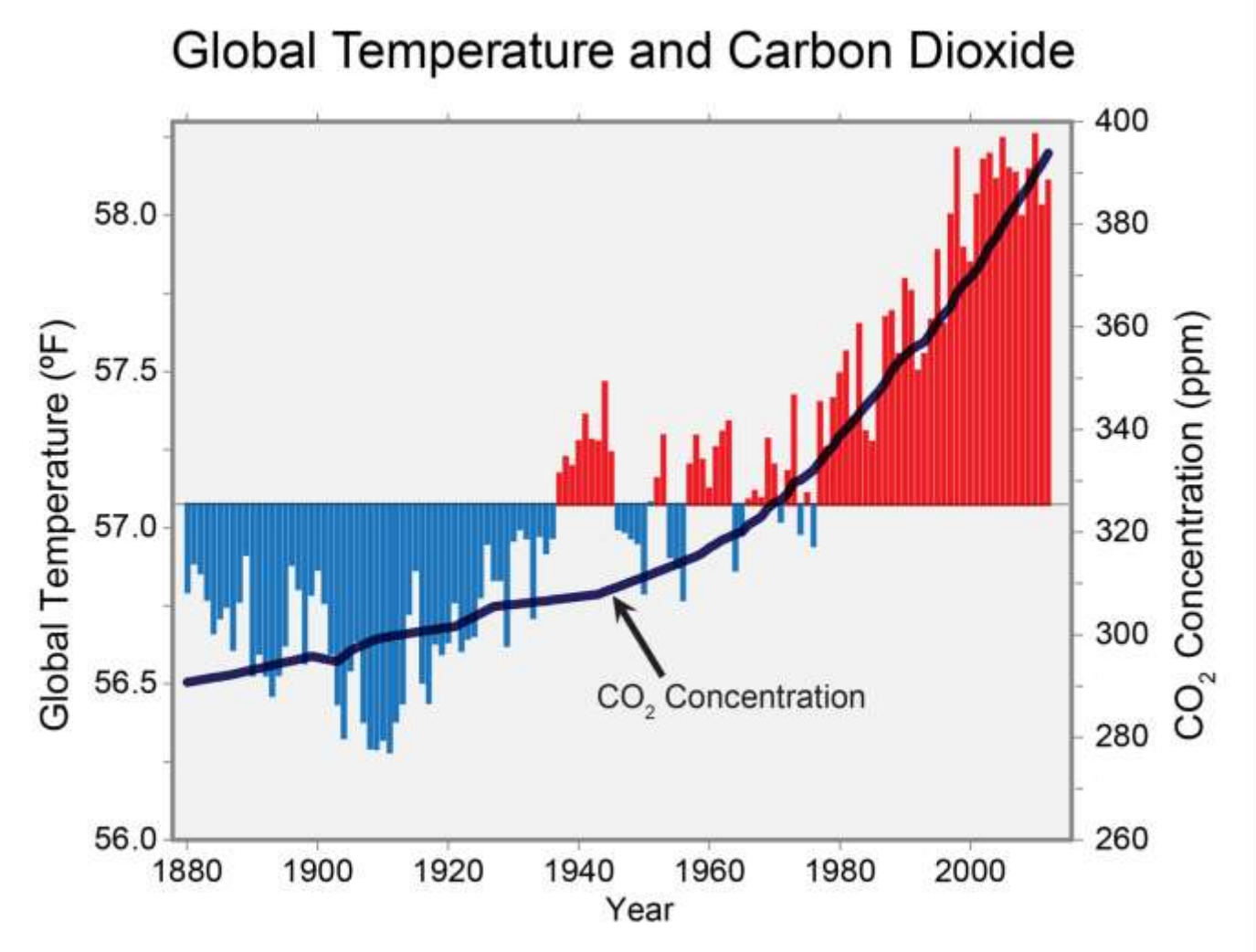
INCREASING REQUIREMENTS

Adapted from AGU EOS 2012 Bates and Privette

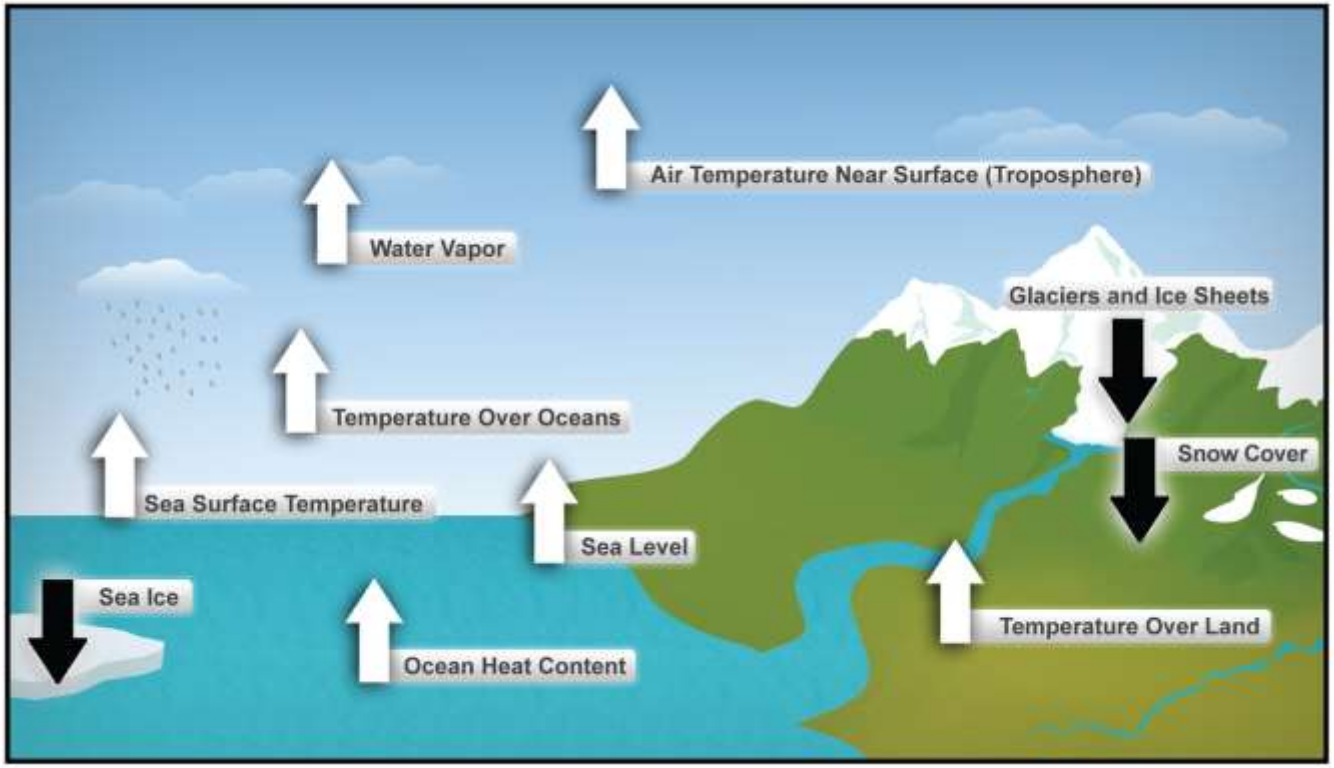
One Slide: Yes, it is getting warmer



Many Lines of Evidence Indicate Current Warming is Human-Caused



Ten indicators of a warmer world



An 11th Indicator: Changing Extremes

Climate Change and Extremes

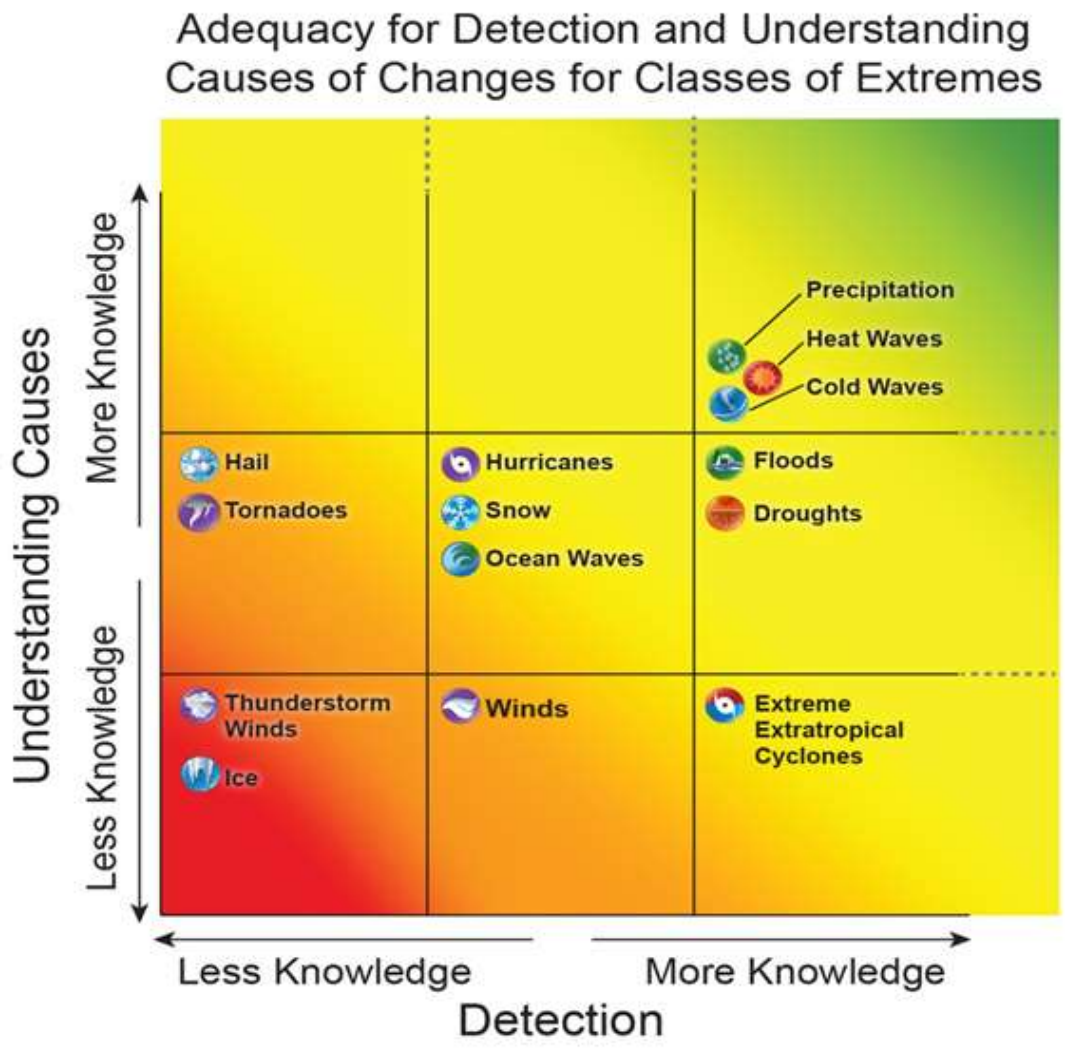
Climate change does not ‘cause’ extreme events. It can change the odds of an extreme event occurring.

Extreme events are born from a set of ingredients.

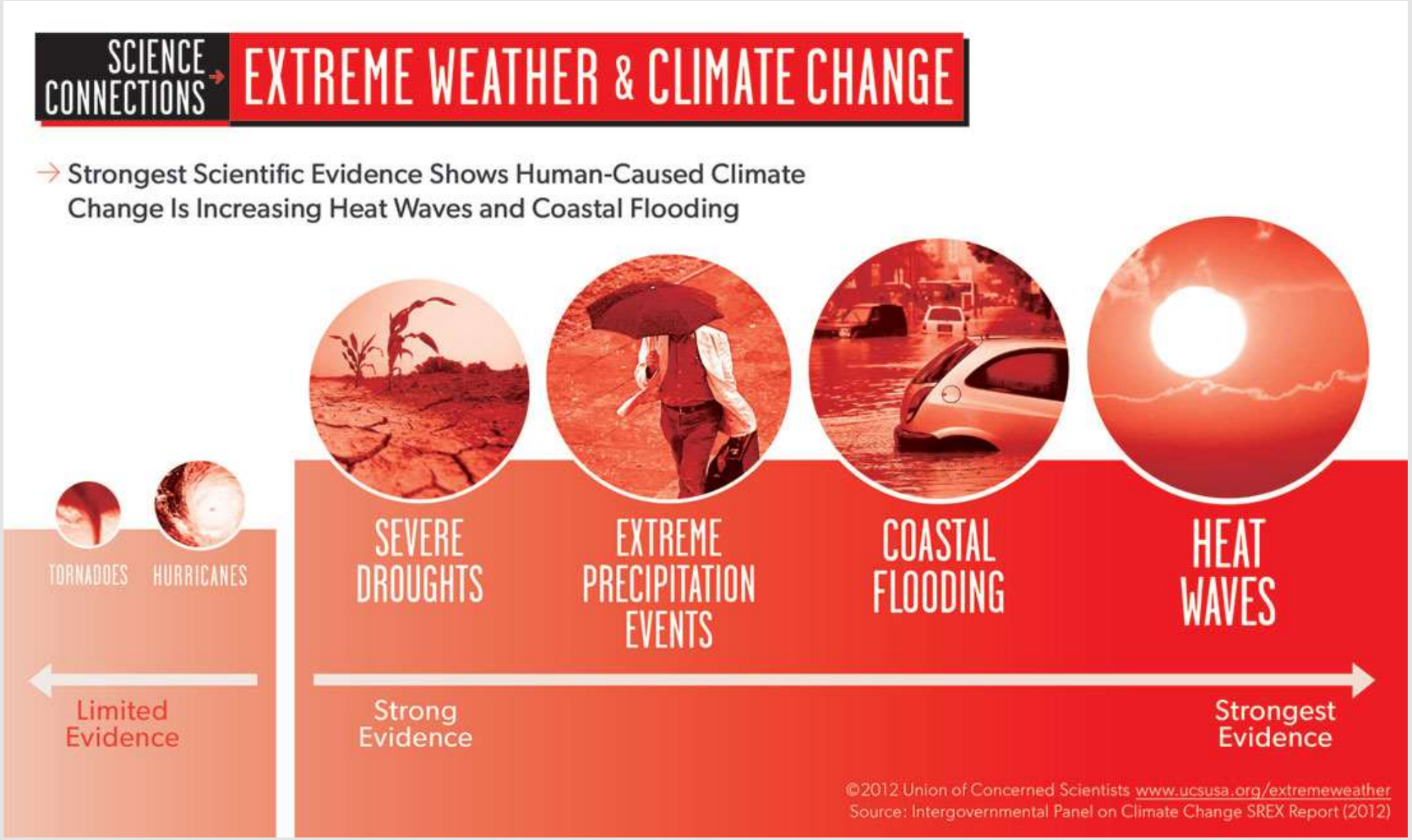
Hurricanes, for example:

- Warm ocean water (temperature of 80°F or greater)
- Form between 5° to 20° latitude
- Lots of moisture in the air
- Light winds throughout the troposphere (the lowest layer of Earth’s atmosphere and where all weather occurs)
- An area of low pressure

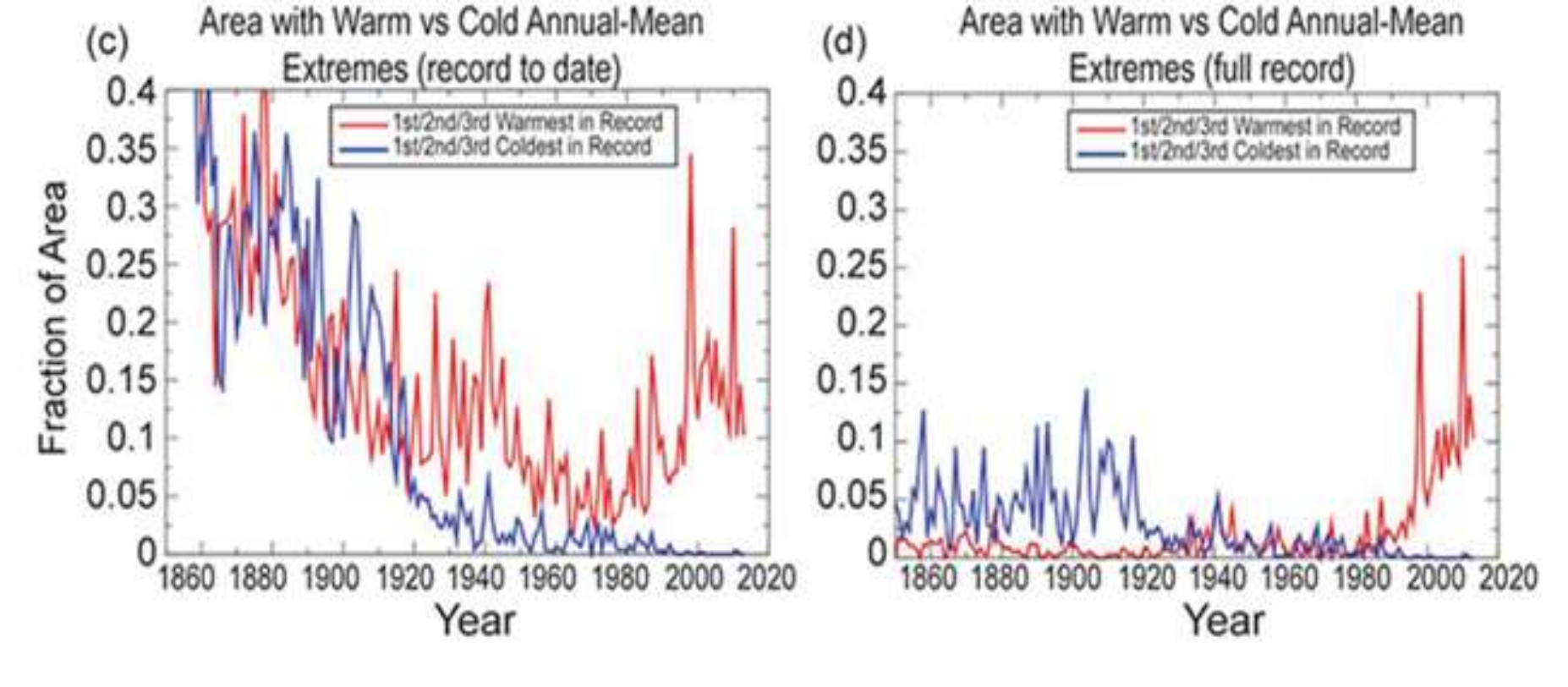
Knowledge Varies by Event Type



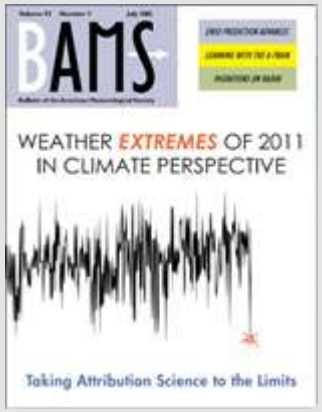
What is changing?

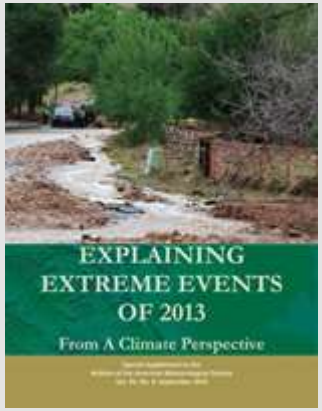


Climate Change and Extreme Temperatures



Explaining Extremes From a Climate Perspective





Looking for the fingerprints of climate change on individual extreme events.

- Why do event attribution?
- What do the results mean?